**CREDIT RISK DATA ANALYSIS**

In a loan, the borrower initially borrows a sum of money(principal) and is obligated to pay an equal amount of money along with interest to the lender at a later time.

Problem statement: When a loan is given out by the lender, it also causes a credit risk. A credit risk is the risk of default on a loan that may arise from a debtor failing to make necessary payments. The risk to the lender comprises of lost principal and interest, disorder in cash flows, and an increase in collection charges.

To determine the credit risk accompanying a loan portfolio, it would be worthwhile to forecast whether a borrower will default. Consequently, actions can be taken to mitigate possible loss to the creditor.

In this project, Probability of Default model has been developed with the help of past credit information about existing borrowers and their behavior with the bank to calculate and measure the risk of each borrower.

Probability of default is a monetary term labeling the likelihood of a default over a certain time horizon. It offers an estimate of the likelihood that a debtor will be unable to meet his/her loan obligations. PD is used in an assortment of credit analyses and risk management background. The probability of default is an estimation of the chances that the default event will happen.

Based on the credit risk the lender can take the following steps for high-risk applicants:

* Lessening credit/services if the risk level is too high
* Allocating a lower starting credit limit on a credit card or line of credit
* Charging greater interest rate on a loan

**About the data**

Our dataset was sourced from [*https://www.lendingclub.com/info/download-data.action*](https://www.lendingclub.com/info/download-data.action). This is Lending Club’s complete loan data for all loans issued through 2007 - 2011. Lending Club is the world’s largest online credit marketplace which facilitates in personal loans, business loans, and financing for elective medical procedures. Investors provide the capital to enable many of the loans in exchange for earning interest. Lending club operates fully online with no branch infrastructure.

This dataset contains a total of 42538 rows and 115 columns. Out of these 115 columns over 68 columns were missing values. The includes the data regarding the current loan status and latest payment information.